A Unified Approach to Global Convergence of Trust-Region Methods for Nonsmooth Optimization

J.E. Dennis
Shou-Bai Li
and
R.A. Tapia

September 1989
(revised July 1993)

TR89-5

(to appear in Mathematical Programming)

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **1989**	2. REPORT TYPE	3. DATES COVERED **00-00-1989 to 00-00-1989**

4. TITLE AND SUBTITLE **A Unified Approach to Global Convergence of Trust Region Methods for Nonsmooth Optimization**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Computational and Applied Mathematics Department ,Rice University,6100 Main Street MS 134,Houston,TX,77005-1892**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**41**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

A Unified Approach to Global Convergence of Trust Region Methods for Nonsmooth Optimization*

John E. Dennis Jr., Shou-Bai B. Li and Richard A. Tapia†

Key Words: nonsmooth optimization, trust region methods, global convergence
Abbreviated Title: Global Convergence of Trust Region Methods

Abstract

This paper investigates the global convergence of trust region (TR) methods for solving nonsmooth minimization problems. For a class of nonsmooth objective functions called regular functions, conditions are found on the TR local models that imply three fundamental convergence properties. These conditions are shown to be satisfied by appropriate forms of Fletcher's TR method for solving constrained optimization problems, Powell and Yuan's TR method for solving nonlinear fitting problems, Zhang, Kim and Lasdon's successive linear programming method for solving constrained problems, Duff, Nocedal and Reid's TR method for solving systems of nonlinear equations, and El Hallabi and Tapia's TR method for solving systems of nonlinear equations. Thus our results can be viewed as a unified convergence theory for TR methods for nonsmooth problems.

*Research supported by AFOSR 89-0363, DOE DEFG05-86ER25017 and ARO 9DAAL03-90-G-0093.

†Department of Computational and Applied Mathematics and Center for Research on Parallel Computation, Rice University, Houston, Texas 77251-1892.

1 Introduction

Trust region methods (TR) are an important class of iterative methods for solving nonlinear optimization problems. In an unconstrained minimization problem, a step to a new iterate is obtained by minimizing a local model of the objective function over a restricted region centered about the current iterate. The size of this restricted region depends on how well the local model predicts the behavior of the objective function. This strategy will force convergence of the iterates from an arbitrary starting point to a point which satisfies the first-order necessary conditions. Motivation and a survey of TR methods for the smooth case can be found in Moré (1983), see also Chapter 6 of Dennis and Schnabel (1983) for unconstrained problems. In the past decade, many trust region methods for minimizing a nonsmooth objective function have been proposed and applied to the nonlinear equations problem, the nonlinear fitting problem, and the constrained optimization problem.

In this paper we consider trust region methods for the unconstrained minimization of a nonsmooth function, *i.e.*,

$$\min f(x) \tag{1}$$

where $f : R^n \rightarrow R$ is a nonsmooth function and may represent the objective function of a nonsmooth unconstrained optimization problem or a nonsmooth penalty function for a constrained optimization problem, such as the l_∞ norm or l_1 norm penalty function.

We begin the TR iteration from a starting point x_0 which may not be close to a solution of (1). Let $L_0 = \{x \mid f(x) \leq f(x_0)\}$ be the level set of f at x_0. We build at each iteration a local model $m(x_k, p_k)(s)$ which is an approximation of $f(x_k + s)$ for small s, and $p_k \in R^l$ is an l-dimensional parameter vector which may change from iteration to iteration. For example, p_k might specify model curvature. Next we approximately solve the subproblem $SUB(x_k, p_k; \delta_k)$:

$$\begin{array}{ll} min & m(x_k, p_k)(s) \\ s.t. & \|s\| \leq \delta_k, \end{array}$$

to obtain a trial step s_k that satisfies

$$\|s_k\| \leq \delta_k \tag{2}$$

and

$$m(x_k, p_k)(0) - m(x_k, p_k)(s_k) \geq \tau[m(x_k, p_k)(0) - m(x_k, p_k)(s_k^*)] \geq 0 \tag{3}$$

where s_k^* is a solution of $SUB(x_k, p_k; \delta_k)$, *i.e.*,

$$s_k^* \in argmin\{m(x_k, p_k)(s) \mid \|s\| \leq \delta_k\},$$

and $0 < \tau \leq 1$ is a fixed constant. The positive number δ_k is called the TR radius. We accept the step s_k and set $x_{k+1} = x_k + s_k$ if

$$r_k = \frac{ared_k}{pred_k} > c_0 \tag{4}$$

where c_0 is a fixed constant in $(0, 1)$, and

$$\begin{aligned} ared_k = &\quad f(x_k) - f(x_k + s_k) \quad (actual\ \ reduction) \\ pred_k = &\quad f(x_k) - m(x_k, p_k)(s_k) \quad (predicted\ \ reduction). \end{aligned}$$

Otherwise we repeat this process using a smaller δ_k in $SUB(x_k, p_k; \delta_k)$. This leads us to the following basic TR algorithm:

at the k-th iteration,

STEP 1 : approximately solve the subproblem $SUB(x_k, p_k; \delta_k)$ to obtain s_k satisfying (2) and (3);

STEP 2 : compute r_k according to (4);

STEP 3 : if $r_k \leq c_0$, then set $x_{k+1} = x_k$, $p_{k+1} = p_k$, reduce δ_k by $\rho_0\delta_k$ and go to STEP 1,
otherwise, set $x_{k+1} = x_k + s_k$, update the TR radius δ_k to δ_{k+1}, update p_k to p_{k+1} and go to STEP 1;

where $1 > c_0 > 0$ and $1 > \rho_0 > 0$ are fixed constants.

It is obvious that this is a conceptual TR algorithm since we have omitted details needed to specify a complete procedure, for example, a stopping criterion, an updating rule for δ_k and a numerical method for determining s_k in STEP 1.

An approximate solution s_k of the TR subproblem $SUB(x_k, p_k; \delta_k)$ is required to satisfy criterion (3). This implies that the step s_k attains at least a fixed fraction τ of the optimal decrease that can be obtained from the TR subproblem. The exact solution s_k^* of the TR subproblem appears in (3) only for comparative purposes. We expect that in general it will not be necessary to compute the exact solution in STEP 1. In the smooth case, Byrd, Schnabel and Shultz (1988) prove under a mild assumption that the widely used sufficient decrease criterion for an approximate solution of the TR subproblem implies (3). If the TR subproblem can be transformed into a linear programming problem, criterion (3) may be checked using information from the dual problem. There is an advantage of using criterion (3) in the nonsmooth case because it does not require gradient information.

This paper focuses on a unified approach to the global convergence of our basic TR algorithm. We will attempt to identify some general assumptions on the objective function and the local model that will allow us to establish the following three fundamental convergence properties of the basic TR algorithm:

1. An iterate x_k is a stationary point of f in (1) if for $\delta_k > 0$, the step $s_k = 0$ is obtained in STEP 1.

2. Reducing δ_k in STEP 3 eventually guarantees $r_k > c_0$ where $1 > c_0 > 0$. Equivalently, if the basic TR algorithm loops infinitely often between STEP 1 and STEP 3 with $x_{k+1} = x_k$, then the current iterate x_k must be a stationary point of f in (1).

3. Any accumulation point of $\{x_k\}$ generated by the basic TR algorithm is a stationary point of f in (1).

In this way we will have obtained a general convergence theory for TR methods. We will then show that we have a useful theory by demonstrating that our theory can handle various TR methods which appear in the literature. The first convergence property is considered in §2. In §2 we also introduce some notation and terminology for nonsmooth functions, and the assumptions that will lead us to a unified global convergence theory. A brief survey of several nonsmooth TR methods is contained in §3. We show that these TR methods satisfy the assumptions introduced in §2. The second and the third convergence properties are considered in §4. In §5, we make some concluding remarks.

2 Assumptions

The convergence analysis presented in this paper is based on some reasonable assumptions on the objective functions and the local models employed in these TR methods. We introduce notation in §2.1, state our assumptions in §2.2 and derive several properties which are related to these assumptions in §2.3 and §2.4.

2.1 Notation and Terminology for Nonsmooth Functions

For our applications, it is reasonable to assume that the objective function $f(x)$ and the local model $m(x,p)(s)$ are always finite, or $f(x) < \infty$ for at least one point x in the level set L_0 and $m(x,p)(s) < \infty$ at $s = 0$. The bulk of the material listed below comes from Clarke (1983) Chapter 2.

Definition 2.1 *A function $f : R^n \to R$ is said to be (locally) Lipschitz near x if for some constants $K > 0$ and $\varepsilon > 0$, f satisfies the Lipschitz condition*

$$| f(x_1) - f(x_2) | \leq K\|x_1 - x_2\|$$

for all x_1 and x_2 in the ε-neighborhood $N(x) = \{y | \|y - x\| < \varepsilon\}$ of x, where $\|\cdot\|$ is a given norm on R^n. The function f is said to be (locally) Lipschitz on a set U if it is (locally) Lipschitz near every point $x \in U$.

Definition 2.2 *The generalized directional derivative of f at x in the direction $d \in R^n$ is*

$$f^\circ(x;d) \triangleq \limsup_{y \to x,\, t \downarrow 0} \frac{f(y+td)-f(y)}{t}.$$

Lemma 2.3 *Let f be Lipschitz with constant K near x. Then for every $d \in R^n$, $f^\circ(x;d)$ exists and*

$$| f^\circ(x;d) | \leq K\|d\|.$$

Definition 2.4 *The generalized gradient of f at x is the set*

$$\partial f(x) \triangleq \{g \in R^n \mid g^T d \leq f^\circ(x;d),\ \forall d \in R^n\}.$$

Definition 2.5 *A point x is said to be a stationary point of f if $0 \in \partial f(x)$.*

Theorem 2.6 (first-order necessary condition) *Let f be Lipschitz near x. If f attains a local minimum at x, then $0 \in \partial f(x)$,* i.e., *x is a stationary point of f.*

Definition 2.7 *A function f that is Lipschitz near x is said to be regular at x if the one-sided directional derivative*

$$f'(x;d) \triangleq \lim_{t \downarrow 0} \frac{f(x+td)-f(x)}{t}$$

exists for all directions $d \in R^n$ and

$$f'(x;d) = f^\circ(x;d).$$

The function f is said to be regular on a set U if it is regular at every point of the set U.

Since the nonsmooth functions discussed in this paper are always assumed to be locally Lipschitz, we defined regularity in Definition 2.7 only for locally

Lipschitz functions. This is slightly different than the definition in Clarke's book. Convex functions defined on open convex sets are locally Lipschitz as is demonstrated by the following theorem. Thus, by Lemma 2.3, a convex function defined on an open convex set always has a generalized directional derivative. From convex analysis (for example, see Theorem 23.1 on page 213 of Rockafellar (1970)), a convex function always has a one-sided directional derivative. The following theorem says that these two derivatives coincide for a convex function.

Theorem 2.8 *Let U be an open convex set in R^n and let f be a convex function on U with $f(\bar{x}) < +\infty$ for some $\bar{x} \in U$. Then f is Lipschitz near any point x in U and $f^\circ(x;d)$ coincides with $f'(x;d)$ for all d in R^n. Thus f is regular on U.*

For composite functions, we have a similar result.

Corollary 2.9 *Let U be an open convex set in R^n and let f be a convex function on U with $f(\bar{y}) < +\infty$ for some $\bar{y} \in U$. Let $g : V \subset R^m \to U \subset R^n$ be continuously differentiable on V. Then the composite function $c(x) = f(g(x))$ is Lipschitz near any point x in V and $c^\circ(x;d)$ coincides with $c'(x;d)$ for all d in R^m. Thus the composite function is regular on V.*

While the following characterization of stationary points does not appear explicitly in Clarke's book, it is well known and quite useful. In a minimization problem, the following lemma says that a point is a stationary point of a locally Lipschitz function if and only if, at this point, there do not exist any descent directions for the function.

Lemma 2.10 *Assume that f is Lipschitz near x. Then $0 \in \partial f(x)$ if and only if $f^\circ(x;d) \geq 0$ for all d in R^n.*

Proof. The existence of $f^\circ(x;d)$ is guaranteed by by Lemma 2.3. The proof now follows in a direct manner. □

2.2 Assumptions

Let $L_0 = \{x | f(x) \leq f(x_0)\}$ denote the level set of f in problem (1) at the starting point x_0, and $m(x,p)'(0;d)$ (or $m(x,p)^\circ(0;d)$) be the one-sided (or generalized) directional derivative of $m(x,p)(s)$ with respect to s at $s = 0$ along the direction $d \in R^n$ for $p \in P \subset R^l$. We employ the following basic assumptions on the objective function and the TR local models.

Assumption 2.1 : The objective function f is regular on L_0.

Assumption 2.2 : For every $(x,p) \in L_0 \times P$, the local model $m(x,p)(s)$ is regular with respect to $s \in R^n$.

Assumption 2.3 : For $(x,p) \in L_0 \times P$, the local model $m(x,p)(s)$ satisfies

$$m(x,p)(0) = f(x)$$

and

$$m(x,p)^\circ(0;d) = f^\circ(x;d), \quad \forall d \in R^n \sim \{0\}.$$

Assumption 2.4 : For every $s \in R^n$, the local model $m(x,p)(s)$ is continuous in (x,p).

Assumption 2.5 : The set P of possible parameter vectors is bounded.

A general theory cannot accomodate the delicate details of all applications. In this sense Assumption 2.5 is restrictive. For example, Powell (1984) and Yuan (1983) developed convergence theories without assuming bounded parameters. However, as we shall see, in many applications Assumption 2.5 is realistic.

When the local model is convex in s, since convexity implies regularity by Theorem 2.8, Assumption 2.2 will be satisfied.

Assumption 2.3 requires every local model to be at least a first-order approximation to the objective function. We observe that no uniformity in (x,p) is involved in this assumption. In general, we are free in each TR iteration to choose the parameter vector p_k provided that the local models always keep first-order approximation to the objective function and the associated parameter vector p_k remains in the bounded set P. In order to verify Assumption 2.3 for a TR method, we will need to know something about the structure of the local model.

The locally Lipschitz level of generalization for nonsmooth objective functions makes it possible to analyze TR methods involving some nonsmooth composite functions other than polyhedral norms. For example, the function

$$c(x) = max(0, min(x_1, x_2))$$

where $x = (x_1, x_2)$, is locally Lipschitz but is not a polyhedral norm. The TR method for solving the optimization problem

$$min\ c(F(x))$$

where $F(x) = (f_1(x_1, x_2), f_2(x_1, x_2))$ is a smooth function, may use the local model, with first-order approximation in F,

$$m(x)(s) = c(F(x) + F'(x)s).$$

The resulting TR subproblem can be converted to a linear programming problem and can be approximately solved at each iteration by a simplex method or an interior point method. See Li (1989) Chapter 5 for details of the convergence analysis. In this paper we deal only with regular objective functions and we do not consider methods for solving the local models.

2.3 Preventing False Termination of the TR Algorithm

The TR subproblem must be posed so as to avoid false termination, *i.e.*, if s_k obtained from the basic TR algorithm in <u>STEP 1</u> with $\delta_k > 0$ happens to be zero, then the algorithm has converged, *i.e.*, x_k is a stationary point of f in (1). We first show that obtaining $s = 0$ from the basic TR algorithm in <u>STEP 1</u> with $\delta > 0$ is equivalent to the fact that $s^* = 0$ solves the TR subproblem $SUB(x, p; \delta)$ with $\delta > 0$, *i.e.*, if the approximate solution satifying the two conditions (2) and (3) is zero, then so is the exact solution.

Lemma 2.11 *Let $\delta > 0$ be the TR radius in the subproblem $SUB(x, p; \delta)$. If the step s obtained from the basic TR algorithm in* <u>*STEP 1*</u> *is equal to zero, then $s^* = 0$ solves $SUB(x, p; \delta)$.*

Proof. Recall that the step s obtained from the basic TR algorithm in <u>STEP 1</u> satisfies (2)

$$\|s\| \leq \delta$$

and (3)

$$m(x,p)(0) - m(x,p)(s) \geq \tau[m(x,p)(0) - m(x,p)(s^*)] \geq 0. \tag{5}$$

If $s = 0$, then $m(x,p)(0) - m(x,p)(s^*) = 0$, which implies $s^* = 0$ solves the subproblem. □

We will now show that Assumptions 2.1 through 2.3 are sufficient for preventing false termination.

Lemma 2.12 *Under Assumptions 2.1 through 2.3, if $s^* = 0$ solves the subproblem $SUB(x, p; \delta)$ with $\delta > 0$, then x is a stationary point of f.*

Proof. Let x and p be fixed. If $s^* = 0$ solves $SUB(x, p; \delta)$ with $\delta > 0$, then $s^* = 0$ is a local minimizer of $m(x,p)(s)$. From the first-order necessary condition (Theorem 2.6), $s^* = 0$ is a stationary point of $m(x,p)(s)$ considered as a function of s. Since $m(x,p)(s)$ is Lipschitz in s near 0, from Lemma 2.10, we have

$$m(x,p)^\circ(0; d) \geq 0, \quad \forall d \in R^n.$$

By Assumption 2.3,

$$m(x,p)^\circ(0; d) = f^\circ(x; d), \quad \forall d \in R^n \sim \{0\}.$$

So

$$f^\circ(x; d) \geq 0, \quad \forall d \in R^n.$$

From Lemma 2.10, $0 \in \partial f(x)$ and x is therefore a stationary point of f. □

Thus under the assumptions of Lemma 2.12, if x is not a stationary point of the objective function f in (1), then any solution s^* of $SUB(x, p; \delta)$ with $\delta > 0$ will not be zero. Therefore from Lemma 2.11, no trial step s obtained in <u>STEP 1</u> will be zero. We will use this result in our later analysis.

2.4 Conditions Equivalent to First-Order Approximation

Assumption 2.3 requires that for regular functions, the one-sided directional derivatives of the objective function f in (1) and the local model $m(x,p)(s)$ must coincide. In many nonsmooth cases, it is more convenient to check if

$$\theta(x,p)(s) \triangleq \frac{f(x+s) - m(x,p)(s)}{\|s\|}$$

converges to zero as $\|s\|$ converges to zero, or if

$$\theta(x,p)(td) = \frac{f(x+td) - m(x,p)(td)}{|t|\ \|d\|}$$

converges to zero for all $d \in R^n \sim \{0\}$ as $t \downarrow 0$. The next lemma shows that these three conditions are equivalent for regular functions.

Lemma 2.13 *Under Assumptions 2.1 and 2.2, for $(x,p) \in L_0 \times P$, the following three conditions are equivalent:*

1. *Assumption 2.3 :*
$$m(x,p)(0) = f(x),$$
$$m(x,p)^\circ(0;d) = f^\circ(x;d), \quad \forall d \in R^n \sim \{0\};$$

2. $$\lim_{\|s\| \to 0} \theta(x,p)(s) = 0;$$

3. $$\lim_{t \downarrow 0} \theta(x,p)(td) = 0$$
for all $d \neq 0$ in R^n.

Furthermore, any one of these conditions implies that, if the step s obtained from the basic TR algorithm in STEP 1 with $\delta > 0$ in $SUB(x,p;\delta)$ is equal to zero, then x is a stationary point of f.

Proof. We first prove that the first condition implies the second condition. Suppose that Assumption 2.3 holds. If the limit of $\theta(x,p)(s)$ is not zero, then there must exist a sequence $\{s_k\}$ and $\varepsilon > 0$ such that $\|s_k\| \to 0$ and

$$|\theta(x,p)(s_k)| = \frac{|f(x+s_k) - m(x,p)(s_k)|}{\|s_k\|} \geq \varepsilon.$$

Let $t_k = \|s_k\|$, $d_k = s_k/\|s_k\|$. Since $m(x,p)(0) = f(x)$, we can write

$$\theta(x,p)(s_k) = \frac{f(x+t_kd_k) - f(x)}{t_k} - \frac{m(x,p)(0+t_kd_k) - m(x,p)(0)}{t_k}.$$

Since $\|d_k\| = 1$, there must exist a subsequence $\{d_{k_i}\}$ and d_* such that $\|d_{k_i}\| = \|d_*\| = 1$, $d_{k_i} \to d_*$ as $i \to \infty$, and $|\theta(x,p)(s_{k_i})| \geq \varepsilon$. Let $\{K'\}$ denote the index set $\{k_i\}$ of the above subsequence. The first term of $\theta(x,p)(s_{k_i})$ can be written

$$\frac{f(x+t_kd_k) - f(x)}{t_k} = \frac{f(x+t_kd_k) - f(x+t_kd_*)}{t_k} + \frac{f(x+t_kd_*) - f(x)}{t_k}$$

for $k \in \{K'\}$. Since f is Lipschitz near x from Definition 2.7, we have

$$|\frac{f(x+t_{k_i}d_{k_i}) - f(x+t_{k_i}d_*)}{t_{k_i}}| \leq K\|d_{k_i} - d_*\|$$

where K is the Lipschitz constant of f near x, and

$$\lim_{i\to\infty} \frac{f(x+t_{k_i}d_*) - f(x)}{t_{k_i}} = f^\circ(x;d_*).$$

Thus the first term of $\theta(x,p)(s_{k_i})$ converges to $f^\circ(x;d_*)$ as $i \to \infty$. Similarly the second term of $\theta(x,p)(s_{k_i})$ converges to $m(x,p)^\circ(0;d_*)$ as $i \to \infty$. By Assumption 2.3, $m(x,p)^\circ(0;d_*) = f^\circ(x;d_*)$. Thus the limit of $\theta(x,p)(s_{k_i})$ exists and is equal to 0 as $i \to \infty$, which contradicts $|\theta(x,p)(s_{k_i})| \geq \varepsilon$. Therefore the limit of $\theta(x,p)(s)$ must be zero as $\|s\|$ converges to zero.

It is obvious that the second condition implies the third condition. Finally we will show that the third condition implies the first condition. Suppose that

$$\lim_{t\downarrow 0} \theta(x,p)(td) = \lim_{t\downarrow 0} \frac{f(x+td) - m(x,p)(td)}{t\ \|d\|} = 0$$

for all d in $R^n \sim \{0\}$. Recalling the definition of $\theta(x,p)(s)$, we have

$$f(x+td) = m(x,p)(td) + \theta(x,p)(td)\ t\ \|d\|$$

where

$$\lim_{t\downarrow 0} \theta(x,p)(td) = 0$$

for all d in $R^n \sim \{0\}$. Since $f(x+td)$ and $m(x,p)(td)$ as functions of t are continuous at $t=0$, by letting $t \downarrow 0$, we obtain $f(x) = m(x,p)(0)$. It also follows from the above expression that

$$\begin{aligned} f^\circ(x;d) &= \lim_{t\downarrow 0} \frac{f(x+td)-f(x)}{t} \\ &= \lim_{t\downarrow 0}[\frac{m(x,p)(td)-m(x,p)(0)}{t} + \theta(x,p)(td)\ \|d\|] \\ &= m(x,p)^\circ(0;d) \end{aligned}$$

for all d in $R^n \sim \{0\}$.

From Lemma 2.11 and Lemma 2.12, the first-order approximation in Assumption 2.3 or any of the equivalent conditions also guarantees that x is a stationary point of f in (1) when the step s obtained from the basic TR algorithm in STEP 1 with $\delta > 0$ in $SUB(x,p;\delta)$ is equal to zero. □

3 Case Studies

As nonsmooth norms and nonsmooth penalty or merit functions are widely employed in both constrained and unconstrained optimization problems, nonsmooth TR methods and their convergence analysis have become an active research area in recent years. We will introduce some nonsmooth TR methods in this section and show that the assumptions listed in Section 2 are reasonable in that they apply to these TR methods. Therefore any convergence theory which follows from these assumptions can be viewed as a unified approach to global convergence for TR methods.

We let ∇ denote the gradient operator and ∇^2 the Hessian operator of a functional defined in R^n. We also use the notation $\nabla c(x)$ to denote $c'(x)^T$, the transpose of the Jacobian matrix at x for a function $c: R^n \rightarrow R^m$.

3.1 The Smooth Problem

The smooth problem

$$\min f(x),$$

where $f : R^n \to R$ is assumed to be continuously differentiable, can be solved by the basic TR algorithm with the local model

$$m(x_k, B_k)(s) = f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T B_k s,$$

where each B_k in $m(x_k, B_k)(s)$ is assumed to belong to a bounded set P in $R^{n\times n}$. Obvious choices for B_k are $B_k = 0$, or $B_k = \nabla^2 f(x_k)$ when $f \in C^2$.

Now we show that the assumptions listed in Section 2 are satisfied for the local model $m(x, B)(s)$. The continuous differentiability of f implies the regularity of f. Since $m(x, B)(s)$ is linear or quadratic in s for every (x, B), the local model is regular in s and Assumption 2.2 is satisfied. Since

$$f(x) = m(x, B)(0),$$

$$f^\circ(x; s) = \nabla f(x)^T s,$$

and

$$m(x, B)^\circ(0; s) = \nabla f(x)^T s,$$

for every $x \in R^n$ and $B \in R^{n\times n}$, Assumption 2.3 holds. Assumption 2.4 holds because continuity in (x, B) of $m(\cdot,\cdot)(s)$ follows from the continuity of $f(\cdot)$ and $\nabla f(\cdot)$. Assumption 2.5 holds as long as the matrix B_k is chosen from a bounded set $P \subset R^{n\times n}$.

Therefore any convergence analysis based on our assumptions will apply to the smooth objective function f and the local model $m(x, B)(s)$ with any norm including a nonsmooth norm employed in the TR subproblem.

Powell (1984) obtained a convergence theorem, which is similar to Theorem 3.2 in Section 3.3, for TR methods for smooth unconstrained minimization under weaker assumptions on the second derivative approximation B_k.

3.2 Fletcher's TR Method

Fletcher (1982), (1984), (1987) suggested a TR method to solve nonsmooth unconstrained problems of the form

$$\min \phi(x) \triangleq f(x) + h(c(x)) \tag{6}$$

where $f : R^n \to R$ and $c : R^n \to R^m$ are twice continuously differentiable, and $h : R^m \to R$ is the polyhedral convex function

$$h(x) = \max_i (h_i^T x + b_i) \tag{7}$$

where the vectors h_i and the scalars b_i are given. The l_1 and l_∞ penalty functions

$$p_1(x) = f(x) + \rho\|c(x)\|_1$$
$$p_\infty(x) = f(x) + \rho\|c(x)\|_\infty$$

with penalty parameter ρ are often used in solving the smooth equality constrained problem

$$\begin{array}{rc} min & f(x) \\ s.t. & c(x) = 0, \end{array}$$

and can be written in the form (6).

The basic model algorithm and convergence analysis of Fletcher's TR method were given in Fletcher (1982). This material is also reviewed in the paper Fletcher (1984) and in the book Fletcher (1987).

The basic local model employed in Fletcher's TR algorithm is

$$\begin{aligned} &m(x_k, \lambda^k)(s) = \\ &\qquad f(x_k) + \nabla f(x_k)^T s + \tfrac{1}{2} s^T B(x_k, \lambda^k) s + h(c(x_k) + \nabla c(x_k)^T s), \end{aligned} \tag{8}$$

where

$$B(x_k, \lambda^k) = \nabla^2 f(x_k) + \sum_{i=1}^{m} \lambda_i^k \nabla^2 c_i(x_k)$$

and λ_i^k is the i-th component of the multiplier λ^k associated with the previous TR subproblem.

The step in Fletcher's algorithm is accepted if

$$r_k = \frac{\phi(x_k) - \phi(x_k + s_k)}{\phi(x_k) - m(x_k, \lambda^k)(s_k)} > c_0 = 0.$$

A global solution of the TR subproblem in STEP 1 is required in Fletcher's algorithm. The following global convergence theorem was given by Fletcher.

Theorem 3.1 (Fletcher (1982), (1987)) *Let the sequence $\{x_k\}$ generated by the TR algorithm be contained in a bounded (convex) set D and let f, c be C^2 functions whose second derivative matrices are bounded on D. Then there exists an accumulation point x^∞ of the iteration sequence such that x^∞ is a stationary point of ϕ.*

We will show that the objective function $\phi(x)$ given by (6) is regular and Assumptions 2.2 through 2.5 hold for the local model $m(x_k, \lambda_k)(s)$ under the assumptions made by Fletcher in Theorem 3.1.

Since h is convex, by Corollary 2.9, $\phi(x)$ given by (6) is regular on D and $m(x, \lambda)(s)$ is regular in s for every $(x, \lambda) \in D \times R^m$, *i.e.*, Assumptions 2.1 and 2.2 hold.

From Lemma 2.13, we may verify Assumption 2.3 by checking if

$$\theta(x, \lambda)(s) = \frac{\phi(x+s) - m(x, \lambda)(s)}{\|s\|}$$

converges to 0 as $\|s\|$ converges to 0 for every $(x, \lambda) \in D \times R^m$. From the Taylor expansion for f at x,

$$f(x+s) = f(x) + \nabla f(x)^T s + \frac{1}{2} s^T \nabla^2 f(\xi) s$$

where $\xi \to x$ as $\|s\| \to 0$. We have

$$\begin{aligned}\theta(x, \lambda) &= \frac{1}{2} s^T [\nabla^2 f(\xi) - \nabla^2 f(x)] \frac{s}{\|s\|} + \frac{h(c(x+s)) - h(c(x) + \nabla c(x)^T s)}{\|s\|} \\ &\quad - \frac{1}{2} \sum_{i=1}^{m} \lambda_i s^T \nabla^2 c_i(x) \frac{s}{\|s\|}.\end{aligned}$$

Since f is twice continuously differentiable on D, the first term of $\theta(x, \lambda)(s)$

$$\frac{1}{2} s^T [\nabla^2 f(\xi) - \nabla^2 f(x)] \frac{s}{\|s\|}$$

converges to 0 as $\|s\| \to 0$. From Theorem 2.8, the convex function h is Lipschitz near x with Lipschitz constant K, which implies that the second term of $\theta(x, \lambda)(s)$

$$\frac{|h(c(x+s)) - h(c(x) + \nabla c(x)^T s)|}{\|s\|} \leq K \frac{\|c(x+s) - c(x) - \nabla c(x)^T s\|}{\|s\|}$$

converges to 0 for every $x \in D$ as $\|s\| \to 0$ because of the Fréchet differentiability of the function c. The third term of $\theta(x, \lambda)(s)$

$$\frac{1}{2} \sum_{i=1}^{m} \lambda_i s^T \nabla^2 c_i(x) \frac{s}{\|s\|}$$

converges to 0 as $\|s\| \to 0$ for every $(x, \lambda) \in D \times R^m$. Therefore Assumption 2.3 holds.

Since f and c are twice continuously differentiable and the convex function h is continuous, the local model $m(x, \lambda)(s)$ is continuous in (x, λ) for every s in R^n, *i.e.*, Assumption 2.4 holds.

Under the above assumptions, Fletcher proved that the multiplier λ^k is uniformly bounded in k. For example, see (1.6) in Fletcher (1984) or Lemma 14.2.1 in Fletcher (1987). This says that the parameter vectors λ^k are chosen from a bounded set, *i.e.*, Assumption 2.5 holds. Fletcher was aware that his convergence theory still holds if the matrices $B(x_k, \lambda^k)$ in (8) are replaced with matrices B_k belonging to a bounded set. See Section 4 of Fletcher (1984) for example. This is also an implication of our unified theory.

Notice that we only deal with the case that $r(x,p)(s) > c_0 > 0$ in this paper, *i.e.*, the TR step is accepted when the ratio is greater than a positive constant c_0. The TR step is accepted in Fletcher's TR algorithm when the ratio $r(x,p)(s) > 0$, *i.e.*, $c_0 = 0$. Fletcher's convergence result in Theorem 3.1 says that there exists an accumulation point of the TR iterates which is a stationary point of the objective function. The convergence result obtained in this paper for the slightly more restrictive acceptance test is the stronger result that any accumulation point of the TR iteration sequence is a stationary point of the objective function. Hence by making a slight modification of Fletcher's algorithm we establish a stronger result.

3.3 Powell and Yuan's TR Method

Powell (1983) considered a TR method to solve

$$\min h(F(x)),$$

where $F(x) = (f_1(x), ..., f_m(x))^T : R^n \rightarrow R^m$ is continuously differentiable and $h : R^m \rightarrow R$ is convex. He used the local model

$$m(x_k, B_k)(s) = h(F(x_k) + F'(x_k)s) + \frac{1}{2}s^T B_k s$$

where B_k is an $n \times n$ symmetric matrix. This general type of nonlinear problem includes the minmax problem

$$\min_x \max_i |f_i(x)| = \|F(x)\|_\infty,$$

the nonlinear l_1 problem

$$\min_x \sum_{i=1}^{m} |f_i(x)| = \|F(x)\|_1,$$

and the nonlinear least-squares problem

$$\min_x \sum_{i=1}^{m} f_i(x)^2 = \|F(x)\|_2^2.$$

Yuan (1983) proved the following global convergence result for Powell's TR method.

Theorem 3.2 (Yuan (1983)) *Assume that F is continuously differentiable and h is convex in R^n. If $h(F(x))$ is bounded below, the sequence $\{x_k\}$ generated by the TR algorithm is bounded, and if the inequality*

$$\| B_k \| \leq c_3 + c_4 k,$$

where c_3, c_4 are positive constants, holds for all k, then there exists an accumulation point of $\{x_k\}$ which is a stationary point of $h(F(x))$.

We only consider the case when B_k is bounded for all k, so that the local model $m(x, B)(s)$ satisfies Assumption 2.5. Since h is convex and F is continuously differentiable, by Corollary 2.9, the composite function $h(F(x))$ is regular on R^n and $h(F(x) + F'(x)s)$ is regular in s for every $x \in R^n$. Since the second term of $m(x, B)(s)$ is differentiable in s for every $B \in R^{n\times n}$, the local model $m(x, B)(s)$ is regular in s for every $(x, B) \in (R^n \times R^{n\times n})$, *i.e.*, Assumptions 2.1 and 2.2 hold.

We verify Assumption 2.3 by considering the second equivalent form in Lemma 2.13. From the Fréchet differentiability of F, we have

$$F(x + s) = F(x) + F'(x)s + \sigma(x, s)\|s\|$$

where

$$\sigma(x, s) \triangleq \frac{F(x + s) - F(x) - F'(x)s}{\|s\|}$$

and

$$\lim_{\|s\|\to 0} \|\sigma(x, s)\| = 0$$

for every $x \in R^n$. Thus,

$$\begin{aligned} &h(F(x + s)) - m(x, B)(s) = \\ &\qquad h(F(x) + F'(x)s + \sigma(x, s)\|s\|) - h(F(x) + F'(x)s) - \frac{1}{2}s^T Bs. \end{aligned}$$

The convexity of h implies from Theorem 2.8 that h is Lipschitz near $F(x) + F'(x)s$ with constant K, so that

$$|\theta(x, B)(s)| = \frac{|h(F(x + s)) - m(x, B)(s)|}{\|s\|} \le K\|\sigma(x, s)\| + \frac{1}{2}\|s\|\|B\|$$

for every $(x, B) \in R^n \times R^{n\times n}$. Therefore

$$\lim_{\|s\|\to 0} \theta(x, B)(s) = 0$$

for every $(x, B) \in R^n \times R^{n\times n}$. By Lemma 2.13, Assumption 2.3 holds.

The local model $m(x, B)(s)$ is continuous in (x, B) for every $s \in R^n$ because of the continuous differentiability of F and the continuity of the convex function h. Thus Assumption 2.4 holds.

We have demonstrated that Powell and Yuan's TR method satisfies Assumption 2.1 through Assumption 2.5 under the assumptions of Theorem 3.2 and the additional assumption that $\{B_k\}$ is bounded. This allows us to use Theorem 4.3 to prove that any accumulation point of the TR iterates is a stationary point of the objective function $h(F(x))$. Theorem 3.2 says that there exists an accumulation point of the TR iterates which is a stationary point of the objective function $h(F(x))$. Thus we make a stronger assumption and derive a stronger conclusion.

3.4 Zhang, Kim and Lasdon's Successive Linear Programming Method

Zhang, Kim and Lasdon (1985) suggested a TR algorithm for solving the constrained problem

$$\begin{array}{rll} min & h_0(x) & \\ s.t. & h_i(x) = 0, & i = 1, ..., k, \\ & h_j(x) \le 0, & j = k+1, ..., m, \end{array}$$

where $h_i : R^n \to R$, $i = 0, 1, ..., m$, are continuously differentiable. They solve the constrained problem by minimizing an l_1 penalty function

$$f(x) = h_0(x) + \sum_{i=1}^{k} w_i \mid h_i(x) \mid + \sum_{i=k+1}^{m} w_i \;\; max(0, h_i(x)),$$

where $w_i, i = 1, ..., m$, are penalty parameters. The local model

$$\begin{aligned} m(x)(s) = h_0(x) + \nabla h_0(x)^T s + \sum_{i=1}^{k} w_i \mid h_i(x) + \nabla h_i(x)^T s \mid \; + \\ \sum_{i=k+1}^{m} w_i \;\; max(0, h_i(x) + \nabla h_i(x)^T s) \end{aligned}$$

is employed in the TR subproblem

$$\begin{array}{rl} min & m(x)(s) \\ s.t. & \|s\|_\infty \le \delta. \end{array}$$

This subproblem can be transformed into a linear programming (LP) problem. Thus they compute a stationary point of the penalty function $f(x)$ by solving a sequence of LPs; hence their algorithm can be viewed as a successive linear programming (SLP) method. They obtained the following global convergence theorem.

Theorem 3.3 (Zhang, Kim and Lasdon (1985)) *If h_i, $i = 0, 1, ..., m$, are continuously differentiable and the level set of $f(x)$ at the initial point of the algorithm is bounded, then the sequence $\{x_k\}$ of TR iterates has accumulation points, and every accumulation point of $\{x_k\}$ is a stationary point of $f(x)$. Furthermore, if an accumulation point is feasible for the constrained problem, then it is a Kuhn-Tucker point of the constrained problem.*

We now show that this TR method satisfies Assumptions 2.1 through Assumption 2.4 under the assumptions made by Zhang, Kim and Lasdon (1985). Assumption 2.5 is not applicable in this case because no parameter vector is involved in the local model $m(x)(s)$. Here the penalty parameters w_i are considered to be constants and are not changed iteration by iteration in the TR method since they appear in both the objective function and the local models. Since every $h_i(x)$ is continuously differentiable and the functions $|t|$, $max(0, t)$ are convex, by Corollary 2.9, the composite function $f(x)$ is regular in R^n. Also $m(x)(s)$ is convex in s for every $x \in R^n$ because of the convexity of the functions $|t|$ and $max(0, t)$. Hence Assumptions 2.1 and 2.2 hold by Corollary 2.9. The continuous differentiability of h_i and the continuity of the functions $|t|$ and $max(0, t)$ imply that $m(x)(s)$ is continuous in x for all $s \in R^n$, *i.e.*, Assumption 2.4 holds.

We need to show that Assumption 2.3 holds. From the Fréchet differentiability of h_i, we have

$$h_i(x + s) = h_i(x) + \nabla h_i(x)^T s + \sigma_i(x, s)\|s\|$$

$$i = 0, 1, \ldots, m$$

where

$$\sigma_i(x, s) \triangleq \frac{h_i(x + s) - h_i(x) - \nabla h_i(x)^T s}{\|s\|} \to 0$$

as $\|s\| \to 0$ for every $x \in R^n$. The inequalities

$$\left|\ |h_i(x+s)| - |h_i(x) + \nabla h_i(x)^T s|\ \right| \leq |\sigma_i(x,s)|\|s\|$$

$$i = 1, \ldots, k$$

and

$$|max(0, h_i(x+s)) - max(0, h_i(x) + \nabla h_i(x)^T s)| \leq |\sigma_i(x,s)|\|s\|$$

$$i = k+1, \ldots, m$$

imply that

$$\frac{|f(x+s) - m(x)(s)|}{\|s\|} \leq |\sigma_0(x,s)| + \sum_{i=1}^{m} w_i |\sigma_i(x,s)|,$$

which shows that

$$\theta(x,s) = \frac{f(x+s) - m(x)(s)}{\|s\|} \to 0$$

as $\|s\| \to 0$ for every $x \in R^n$. By Lemma 2.13, Assumption 2.3 holds.

3.5 Duff, Nocedal and Reid's TR Method

Duff, Nocedal and Reid (1987) suggested a TR method to solve a system of nonlinear equations

$$F(x) = 0$$

where $F : R^n \to R^n$ is continuously differentiable. As a globalization strategy for a locally convergent method, for example, Newton's method or a secant method, they solve the unconstrained nonsmooth problem

$$\min\ f(x) = \|F(x)\|_1,$$

by a TR method with subproblem

$$\begin{array}{ll} \min & m(x_k)(s) = \|F(x_k) + F'(x_k)s\|_1 \\ s.t. & \|s\|_\infty \leq \delta_k. \end{array}$$

They pointed out that in this way one can use LP techniques to solve the TR subproblem in each iteration, and therefore take advantage of any sparsity patterns in the Jacobian $F'(x_k)$ more readily than in an l_2 TR method. Instead of the ratio test in the basic TR algorithm, they employed a sufficient decrease condition

$$\|F(x+s)\|_1 \leq \|F(x)\|_1 - \alpha \, \|F'(x)s\|_1,$$

where $1 > \alpha > 0$ is a fixed constant, to accept or reject the new iterate and used the ratio r_k as a basis for reducing or increasing the TR radius.

Duff, Nocedal and Reid (1987) did not give a convergence result and pointed out that their approach of updating the TR radius is open to improvement. Their method is considered in this paper to be a special case of El Hallabi and Tapia's TR method. Therefore the discussion on the assumptions and the convergence for their method is contained in the next subsection.

3.6 El Hallabi and Tapia's TR Method

El Hallabi and Tapia (1987) analyze an arbitrary norm TR method for solving a system of nonlinear equations

$$F(x) = 0,$$

where $F : R^n \to R^n$ is continuously differentiable. To obtain a solution, they solve the unconstrained nonsmooth problem

$$min \; f(x) = \|F(x)\|,$$

by a TR algorithm as a globalization strategy with the local model

$$m(x)(s) = \|F(x) + F'(x)s\|.$$

Different norms are allowed in the various parts of the subproblem

$$\begin{array}{ll} \min & m(x_k)(s) = \|F(x_k) + F'(x_k)s\|_a \\ s.t. & \|s\|_b \leq \delta_k \end{array}$$

where $\|\cdot\|_a$ and $\|\cdot\|_b$ are any two norms on R^n. In their algorithm, s_k is a solution of the subproblem in STEP 1 , and δ_k is reduced in STEP 3 until a sufficient decrease condition

$$f(x_k + s_k) \le f(x_k) + \alpha\, \gamma(x_k, s_k)$$

is satisfied, where the choice for $\gamma(x; s)$, for example, may be

$$\gamma(x, s) = \|F(x) + F'(x)s\|_a - \|F(x)\|_a = m(x)(s) - m(x)(0).$$

Thus, if $\gamma(x, s) < 0$, then

$$f(x + s) \le f(x) + \alpha\, \gamma(x, s)$$

which is equivalent to

$$\frac{f(x+s) - f(x)}{m(x)(s) - m(x)(0)} \ge \alpha > 0,$$

the standard TR ratio test.

It is worth mentioning that El Hallabi and Tapia (1987) established the inequalities

$$\gamma_3(x, s) \le \gamma_2(x, s) \le \gamma_1(x, s)$$

for the three choices of $\gamma(x, s)$

$$\gamma_1(x, s) = \|F(x) + F'(x)s\| - \|F(x)\|$$

$$\gamma_2(x, s) = f^{\circ}(x; s)$$

$$\gamma_3(x, s) = -\|F'(x)s\|,$$

which means that the decreases $\alpha|\gamma_i(x, s)|$ required by the algorithm in each iteration with these three different choices of $\gamma(x, s)$ satisfy the inequalities

$$\alpha|\gamma_3(x, s)| \ge \alpha|\gamma_2(x, s)| \ge \alpha|\gamma_1(x, s)| > 0.$$

Clearly, γ_3 corresponds to Duff, Nocedal and Reid's test. Also $\gamma_1(x, s)$ is the preferred choice among the three in the sense that it asks the least decrease in each iteration and is most easily satisfied.

El Hallabi and Tapia (1987) proved the following global convergence theorem for their TR algorithm with various upper semi-continuous $\gamma(x, s)$ with respect to (x, s) in the sufficient decrease condition. Their theory handled the choice γ_1, a modified form of γ_2, and could not handle γ_3. They conjectured that the choices γ_2 and γ_3 do not lead to globally convergent algorithms.

Theorem 3.4 (El Hallabi and Tapia (1987)) *If F is continuously differentiable and the level set of f at the initial point is bounded, then any accumulation point of the sequence $\{x_k\}$ generated by the TR algorithm is a stationary point of f.*

Since the different norms in the subproblem do not cause any difficulty in the global convergence we omit the subscripts on the norms in the rest of this section. El Hallabi and Tapia (1987) were aware that global convergence can also be obtained without any difficulty if the function F in the problem is a mapping from R^n to R^m where $m \leq n$. Hence consider $F : R^n \to R^m$ in the rest of this section. We will show that Assumptions 2.1 through 2.4 hold. Thus the convergence analysis given in §4.1 and §4.2 for the basic TR algorithm applies to El Hallabi and Tapia's TR method under their assumptions.

Since the norm is a convex function and F is continuously differentiable, from Corollary 2.9, the composite function f is regular. The convexity of the norms also implies that $m(x)(s)$ is regular in s for every x in R^n, *i.e.*, Assumptions 2.1 and 2.2 hold. It is easy to see that the local model $m(x)(s)$ is continuous in x for every $s \in R^n$, *i.e.*, Assumption 2.4 holds. The Fréchet differentiability of F implies that

$$F(x+s) = F(x) + F'(x)s + \sigma(x, s)\|s\|$$

where

$$\sigma(x, s) \triangleq \frac{F(x+s) - F(x) - F'(x)s}{\|s\|} \to 0$$

as $\|s\| \to 0$ for every $x \in R^n$. Thus

$$|f(x+s) - m(x)(s)| \leq \|F(x+s) - [F(x) + F'(x)s]\| = \|\sigma(x, s)\|\|s\|.$$

It follows that

$$|\theta(x,s)| = \frac{|f(x+s) - m(x)(s)|}{\|s\|} \leq \|\sigma(x,s)\| \to 0$$

as $\|s\| \to 0$ for every $x \in R^n$. From Lemma 2.13, Assumption 2.3 holds. Assumption 2.5 is not applicable in this case because no parameter vector is involved in the local model $m(x)(s)$. We have verified Assumption 2.1 through Assumption 2.4 for the local model under the assumptions made by El Hallabi and Tapia (1987).

4 Convergence Analysis

Recall that the approximate solution s_k of the subproblem $SUB(x_k, p_k; \delta_k)$ obtained in STEP 1 of the basic TR algorithm satisfies conditions (2) and (3). Since an exact solution s_k^* satisfies the above conditions with $\tau = 1$, the convergence results obtained in this section can be applied to a TR iteration that uses the exact solution in STEP 1.

The following theorem says that if the current TR iterate x_k is not a stationary point of f, then there must exist a small TR radius $\hat{\delta}$ and a neighborhood N_k of x_k such that the ratio $r(x,p)(s) > c_0$ for any $x \in N_k$ and $0 < \delta \leq \hat{\delta}$ where s satisfies the two conditions (2) and (3) for the subproblem $SUB(x,p;\delta)$ and $c_0 \in (0,1)$. Thus if the basic TR algorithm loops infinitely often between STEP 1 and STEP 3 with $x_{k+j} = x_k$ and $p_{k+j} = p_k$ for $j \geq 0$, *i.e.*, no such $\hat{\delta}$ and N_k exist, then x_k must be a stationary point of f.

Theorem 4.1 *Let $(\hat{x}, \hat{p}) \in L_0 \times P$ be given and let c_0 be in $(0,1)$. Under Assumptions 2.1 through 2.5, if $\hat{x}$ is not a stationary point of f, then there exist $\hat{\delta} > 0$ and $\varepsilon > 0$, such that, for every x and δ satisfying $\|x - \hat{x}\| < \varepsilon$ and $0 < \delta \leq \hat{\delta}$, we have*

$$r(x,p)(s) = \frac{f(x) - f(x+s)}{f(x) - m(x,p)(s)} > c_0$$

for any s obtained from STEP 1 of the basic TR algorithm for the subproblem $SUB(x,p;\delta)$.

Proof. Since $\hat{x}$ is not a stationary point of f, we know from Lemma 2.10 that there must exist a direction $d \in R^n$ with $\|d\| = 1$ such that $f^\circ(\hat{x}; d) < 0$. Let $\eta = -f^\circ(\hat{x}; d) > 0$. By the definition of the generalized directional derivative,

$$f^\circ(\hat{x}; d) = \limsup_{x \to \hat{x},\, \delta \downarrow 0} \frac{f(x + \delta d) - f(x)}{\delta} = -\eta < 0.$$

Hence there exist $\varepsilon > 0$ and $\hat{\delta} > 0$ such that

$$\frac{f(x + \delta d) - f(x)}{\delta} < -\frac{\eta}{2},$$

i.e.,

$$ared(x, \delta d) = f(x) - f(x + \delta d) > \frac{\eta}{2}\, \delta \tag{9}$$

for any x and δ that satisfies $\|x - \hat{x}\| < \varepsilon$ and $0 < \delta \leq \hat{\delta}$.

From Lemma 2.13, Assumption 2.3 implies that for every $(x, p) \in L_0 \times P$ and, in particular, for every $x \in N(\varepsilon) \triangleq \{x \mid \|x - \hat{x}\| < \varepsilon\}$,

$$f(x + s) = m(x, p)(s) + \theta(x, p)(s)\|s\|$$

where

$$\lim_{\|s\| \to 0} \theta(x, p)(s) = 0.$$

It follows that the actual reduction for any s with $\|s\| \leq \delta$ can be written as

$$\begin{aligned} ared(x, s) &= f(x) - f(x + s) \\ &= f(x) - m(x, p)(s) - \theta(x, p)(s)\|s\| \\ &= pred(x, p)(s) - \theta(x, p)(s)\|s\|, \end{aligned}$$

where $pred(x, p)(s) \triangleq m(x, p)(0) - m(x, p)(s) = f(x) - m(x, p)(s)$ is the predicted reduction. Thus

$$r(x, p)(s) = \frac{ared(x, s)}{pred(x, p)(s)} = 1 - \frac{\theta(x, p)(s)\|s\|}{pred(x, p)(s)}. \tag{10}$$

for every $x \in N(\varepsilon)$. Since $s = \delta d$ is feasible for $SUB(x, p; \delta)$, it follows from (9) that

$$\frac{\eta}{2}\, \delta < ared(x, \delta d) = pred(x, p)(\delta d) - \theta(x, p)(\delta d)\delta \tag{11}$$

for every $x \in N(\varepsilon)$ and $0 < \delta \leq \hat{\delta}$.

Let s^* be an exact solution of $SUB(x, p; \delta)$. Then

$$m(x,p)(s^*) \leq m(x,p)(\delta d)$$

which implies that

$$pred(x,p)(s^*) \geq pred(x,p)(\delta d)$$

for every $(x, p) \in N(\varepsilon)$ and $0 < \delta \leq \hat{\delta}$. From (11), we have

$pred(x,p)(s^*) \geq pred(x,p)(\delta d) > \dfrac{\eta}{2}\, \delta + \theta(x,p)(\delta d)$ for $\delta > 0$ sufficiently small.

Since the right-hand side of the inequality

$$\frac{\delta}{pred(x,p)(s^*)} < \frac{\delta}{\frac{\eta}{2}\delta + \theta(x,p)(\delta d)\delta} = \frac{2}{\eta + 2\theta(x,p)(\delta d)}$$

tends to the constant $2/\eta$ as δ tends to 0, it follows that

$$0 < \frac{\delta}{pred(x,p)(s^*)} \leq \frac{4}{\eta}$$

for small δ. For any s obtained from <u>STEP 1</u> of the basic TR algorithm for the subproblem $SUB(x, p; \delta)$, (10) can be written as

$$r(x,p)(s) = 1 - \theta(x,p)(s)\, \frac{\|s\|}{\delta}\, \frac{\delta}{pred(x,p)(s^*)}\, \frac{f(x) - m(x,p)(s^*)}{f(x) - m(x,p)(s)}. \qquad (12)$$

By conditions (2) and (3) satisfied by s, we have

$$1 \leq \frac{f(x) - m(x,p)(s^*)}{f(x) - m(x,p)(s)} \leq \frac{1}{\tau}$$

and

$$\frac{\|s\|}{\delta} \leq 1.$$

Since the product in the expression (12) converges to 0 as δ converges to 0, it follows that $r(x,p)(s) \to 1$ as $\delta \to 0$. Therefore, under assumptions 2.1

through 2.5, the ratio $r(x,p)(s)$ can be made arbitrarily close to 1 if ε and $\hat{\delta}$ are sufficiently small. Thus we have $r(x,p)(s) > c_0$ for every $x \in N(\varepsilon)$ and $0 < \delta \leq \hat{\delta}$, where $c_0 \in (0,1)$ and s satisfies the two conditions (2) and (3) for $SUB(x,p;\delta)$. □

To put the global convergence theorem on a solid basis, we specify the updating rule for the TR radius δ in STEP 3 of the basic TR algorithm. Let $0 < c_0 < c_1 < 1$ and $0 < \rho_0 < 1 < \rho_1$ be given constants. At the k-th iteration, the basic TR algorithm becomes

STEP 1 approximately solve $SUB(x_k, p_k; \delta_k)$ to obtain s_k satisfying (2) and (3);

STEP 2 compute the ratio r_k according to (4);

STEP 3 if $r_k \leq c_0$, let $x_{k+1} := x_k$, $p_{k+1} := p_k$, $\delta_{k+1} := \rho_0 \delta_k$ and go to STEP 1;

otherwise, let $x_{k+1} := x_k + s_k$, update p_k to p_{k+1} and update δ_k to δ_{k+1}:

$$\delta_{k+1} := \begin{cases} \delta_k & if \quad c_0 < r_k \leq c_1, \\ \rho_1 \delta_k & if \quad c_1 < r_k. \end{cases}$$

In the successful TR iterations where $r_k > c_0$, we set $x_{k+1} = x_k + s_k$ and $s_k \neq 0$. In the unsuccessful TR iterations where $r_k \leq c_0$, we set $x_{k+1} = x_k$ and reduce the TR radius δ_k by $\rho_0 \delta_k$. The feature of the above updating strategy is that the next trial radius δ_{k+1} is not reduced in the successful iterations, but that δ_k approaches 0 for an infinite sequence of unsuccessful steps. Thus the strategy to update the TR radius employed in STEP 3 is

$$\delta_{k+1} = \begin{cases} \rho_0 \delta_k, & if \quad r_k \leq c_0, \\ \delta_k, & if \quad c_0 < r_k \leq c_1, \\ \rho_1 \delta_k, & if \quad c_1 < r_k. \end{cases}$$

The following lemma is needed in the proof of the convergence theorem.

Lemma 4.2 *Let $\{(x_k, p_k)\}$ be a sequence generated by the basic TR algorithm with the updating strategy specified above and let x^* be an accumulation point of $\{x_k\}$ where x^* is not a stationary point of f in (1). Under*

Assumptions 2.1 through 2.5, for every convergent subsequence $x_{k_i} \to x^*$ in $\{x_k\}$ with the ratio $r(x_{k_i}, p_{k_i})(s_{k_i}) > c_0 > 0$, we have

$$\liminf_{i\to\infty} \delta_{k_i} > 0,$$

i.e., *there exists $\beta > 0$ such that $\delta_{k_i} \geq \beta$ for i large enough.*

Proof. Since the iterates remain the same in the unsuccessful iterations, we cancel the repeated unsuccessful iterates in the sequence $\{(x_k, p_k)\}$ so that the sequence $\{(x_k, p_k)\}$ only consists of successful iterates. Let $t_k \geq 0$ denote the number of TR radius reductions to get the successful iterate x_k, *i.e.*, we have t_k unsuccessful iterations before we obtain the successful iterate x_k. The superscript m is used to stand for the m-th radius reduction. According to the update for the next radius,

$$\delta_k^{(m+1)} = \rho_0 \delta_k^{(m)}, \quad m = 0, ..., t_k - 1,$$

$$\delta_k = \delta_k^{(t_k)},$$

and

$$r_k^{(m)} \leq c_0, \quad m = 0, ..., t_k - 1,$$

$$r_k = r_k^{(t_k)} > c_0.$$

In each successful iteration, we have an inital trial radius for the next iteration

$$\delta_{k+1}^{(0)} = \begin{cases} \delta_k, & if \quad c_0 < r_k \leq c_1, \\ \rho_1 \delta_k, & if \quad c_1 < r_k. \end{cases}$$

Hence the inequality

$$\delta_{k+1}^{(0)} \geq \delta_k \tag{13}$$

holds for successful iterations. This reorganization of the sequence $\{(x_k, p_k)\}$ makes no change to the subsequence $\{(x_{k_i}, p_{k_i})\}$ because it only consists of successful iterates. Let $K_1 = \{0, 1, 2, ...\} \supset K_2 = \{k_i\}$.

From Theorem 4.1, there exist a neighborhood $N^* = N(x^*; \varepsilon^*) = \{x | \|x - x^*\| < \varepsilon^*\}$ and a constant $\delta^* > 0$, such that $r(x, p)(s) > c_0$ for every $x \in N^*$ and $0 < \delta \leq \delta^*$, where s satisfies the two conditions (2) and (3) for the subproblem $SUB(x, p; \delta)$. For $k \in K_2$ large enough that $x_k \in N(x^*; \varepsilon^*)$, if the initial trial radius $\delta_k^{(0)} > \delta^*$ for every $k \in K_2$, then $\delta_k^{(0)}$ may need to be reduced t_k times until $r_k = r_k^{(t_k)} > c_0$. Since $r_k^{(t_k-1)} \le c_0$ and $\delta_k^{(t_k-1)} > \delta^*$, the radius

$$\delta_k = \delta_k^{(t_k)} = \rho_0 \delta_k^{(t_k-1)} > \rho_0 \delta^* > 0$$

is bounded below by a positive number for every $k \in K_2$ large enough. Otherwise there exists a thinner subsequence $\{x_k\}$ for $k \in K_3 \subset K_2$ such that the initial trial radius $\delta_k^{(0)} \le \delta^*$ for every $k \in K_3 \subset K_2$ large enough. By Theorem 4.1, $r_k^{(0)} \ge c_0$ for $k \in K_3$. Hence no reduction in the radius is needed, *i.e.*, $t_k = 0$ and $\delta_k = \delta_k^{(0)}$ for every $k \in K_3$. We omit the superscripts on δ_k and r_k for $k \in K_3$ in the rest of the proof.

Suppose

$$\lim_{k \in K_3} \delta_k = 0.$$

From Theorem 4.1 for each constant $c_1 > c_0$, there exist a neighborhood $N^{**} = N(x^*; \varepsilon^{**}) = \{x | \|x - x^*\| < \varepsilon^{**}\} \subset N^*$ and a constant $\delta^{**} > 0$ with $0 < \varepsilon^{**} \le \varepsilon^*$ and $0 < \delta^{**} \le \delta^*$, such that $r(x,p)(s) > c_1$ for every $(x,p) \in N^{**}$ and $0 < \delta \le \delta^{**}$, where s satisfies (2) and (3) for $SUB(x,p;\delta)$. For $k \in K_3$ large enough, we have $x_k \in N^{**}$ and $\delta_k \le \delta^{**}$.

Let $K_3 = \{i_1, i_2, ..., i_j, ...\}$. Consider the TR iterations between $i_{j-1} \in K_3$ and $i_j \in K_3$ for j large enough that

$$x_{i_j} \in N^{**}, \quad \delta_{i_j} \le \delta^{**},$$

i.e., the ones between two successive iterates in the subsequence K_3. If the last iterate x_{i_j-1} before x_{i_j} is not in N^{**}, *i.e.*,

$$\|x_{i_j-1} - x^*\| > \varepsilon^{**},$$

since x_{i_j} is very close to x^* for j large enough that

$$\|x_{i_j} - x^*\| < \frac{1}{2}\varepsilon^{**},$$

then, by (13),

$$\begin{aligned} \delta_{i_j} &= \delta_{i_j}^{(0)} \ge \delta_{i_j-1} \ge \|s_{i_j-1}\| = \|x_{i_j} - x_{i_j-1}\| \\ &\ge \|x_{i_j-1} - x^*\| - \|x^* - x_{i_j}\| \ge \frac{1}{2}\varepsilon^{**}, \end{aligned}$$

which contradicts $\delta_k \to 0$ for $k \in K_3$ as $k \to \infty$. Hence $x_{i_j-1} \in N^{**}$ for j large enough.

Assume that the last $q_j \geq 1$ iterates of x_{i_j} are in N^{**}, *i.e.*,

$$x_{i_j-1}, ..., x_{i_j-q_j} \in N^{**}$$

and $x_{i_j-q_j-1}$ is outside N^{**} for j large enough that

$$\delta_{i_j} \leq \delta^{**}.$$

According to the updating strategy, (13) becomes

$$\delta^{(0)}_{i_j-m+1} \geq \delta_{i_j-m}, \quad m = 1, ..., q_j.$$

Again, if some of these radii are reduced, say $1 \geq m_0 \geq q_j$, which is the largest index among them, *i.e.*,

$$\delta_{i_j-m_0} < \delta^{(0)}_{i_j-m_0},$$

then similarly

$$\delta_{i_j-m_0} \geq \rho_0 \delta^*.$$

Therefore

$$\delta_{i_j} \geq \delta_{i_j-1} = \delta^{(0)}_{i_j-1} \geq ... \geq \delta_{i_j-m_0+1} = \delta^{(0)}_{i_j-m_0+1} \geq \delta_{i_j-m_0} \geq \rho_0 \delta^*,$$

which contradicts $\delta_k \to 0$ for $k \in K_3$ as $k \to \infty$. Thus we only need to consider the worst case

$$\delta_{i_j-m} = \delta^{(0)}_{i_j-m}, \quad m = 1, ..., q_j,$$

which means that the initial trial radius in these iterations is so small that no reduction is necessary. We also omit the superscripts in these iterations. Since $\delta_{i_j} \leq \delta^{**}$ and

$$\delta_{i_j-m+1} \geq \delta_{i_j-m}, \quad m = 1, ..., q_j,$$

we have

$$\delta_{i_j-m} \leq \delta^{**}, \quad m = 0, 1, ..., q_j.$$

From Theorem 4.1,

$$r_{i_j-m} > c_1, \quad m = 0, 1, ..., q_j.$$

According to the updating strategy,

$$\delta_{i_j-m+1} = \rho_1 \delta_{i_j-m}, \quad m = 1, ..., q_j,$$

which implies

$$\delta_{i_j-m} = \frac{\delta_{i_j}}{\rho_1^m}, \quad m = 1, ..., q_j$$

with $\rho_1 > 1$. Since, by (13),

$$\delta_{i_j-q_j} \geq \delta_{i_j-q_j-1},$$

we also have

$$\delta_{i_j-q_j-1} \leq \frac{\delta_{i_j}}{\rho_1^{q_j}}.$$

Thus

$$\sum_{m=0}^{q_j+1} \delta_{i_j-m} \leq \sum_{m=0}^{q_j} \frac{\delta_{i_j}}{\rho_1^m} + \frac{\delta_{i_j}}{\rho_1^{q_j}} \leq 2\delta_{i_j} \sum_{m=0}^{\infty} (\frac{1}{\rho_1})^m = 2\delta_{i_j} \frac{1}{1-\frac{1}{\rho_1}}.$$

However, since $x_{i_j-q_j-1}$ is not in N^{**}, *i.e.*,

$$\|x_{i_j-q_j-1} - x^*\| \geq \varepsilon^{**},$$

and j is large enough that

$$\|x_{i_j} - x^*\| \leq \frac{1}{2}\varepsilon^{**},$$

we have

$$\begin{aligned} \sum_{m=0}^{q_j+1} \delta_{i_j-m} &\geq \sum_{m=0}^{q_j+1} \|s_{i_j-m}\| \geq \| \sum_{m=0}^{q_j+1} s_{i_j-m}\| = \|x_{i_j} - x_{i_j-q_j-1}\| \\ &\geq \|x_{i_j-q_j-1} - x^*\| - \|x_{i_j} - x^*\| \geq \frac{1}{2}\varepsilon^{**}. \end{aligned}$$

Therefore

$$\delta_{i_j} \geq \frac{1}{4}\varepsilon^{**}(1 - \frac{1}{\rho_1}),$$

which contradicts $\delta_k \to 0$ for $k \in K_3$ as $k \to \infty$. This completes the proof. □

We now present our global convergence theorem.

Theorem 4.3 *Under Assumptions 2.1 through 2.5, if x^* is an accumulation point of $\{x_k\}$ generated by the basic TR algorithm, then x^* is a stationary point of f.*

Proof. Suppose that a subseqence $\{x_{k_i}\}$ approaches x^* which is not a stationary point of f. An unsuccessful iterate in the subseqence can be substituted by the same successful iterate because unsuccessful iterates remain the same and make no progress in the basic TR algorithm. If there are some repeated iterates in the subseqence after the substitution, cancel the repeated ones. We still use the same notation $\{x_{k_i}\}$ to represent the substituted and condenced subseqence which only consists of different successful iterates and approaches x^*. Since the parameter vectors $\{p_{k_i}\}$ are bounded by Assumption 2.5, there must exist a thinner subsequence $\{p_k\}$ where $k \in K' = \{k_{i_j}\}$ such that $p_k \rightarrow p^*$ for $k \in K'$. It is worth pointing out that p^* is not necessarily the parameter the modeling technique would associate with x^*. Thus $x_k \rightarrow x^*$ and $p_k \rightarrow p^*$) for $k \in K'$. According to Lemma 4.2, there exists a constant $\beta > 0$ such that

$$\delta_k \geq \beta > 0$$

for $k \in K'$ large enough.

Since $r(x_k, p_k)(s_k) > c_0$ where s_k satisfies conditions (2) and (3) for $SUB(x_k, p_k; \delta_k)$ and $k \in K'$, it follows that

$$f(x_k) - f(x_{k+1}) > c_0[f(x_k) - m(x_k, p_k)(s_k)]. \tag{14}$$

In order to derive a lower bound for the right-hand side of (14), consider the subproblem $SUB(x^*, p^*; \beta)$, and call an exact solution s^*. Since x^* is not a stationary point of f, by Lemma 2.12, $s^* \neq 0$ and

$$f(x^*) - m(x^*, p^*)(s^*) \triangleq \eta^* > 0.$$

The regularity on f implies that f is continuous for every $x \in L_0$. From Assumption 2.4, $m(x,p)(s)$ is continuous in (x,p) for every $s \in R^n$. For $k \in K'$ large enough, therefore, we have

$$f(x_k) - m(x_k, p_k)(s^*) > \frac{\eta^*}{2},$$

where

$$\|s^*\| \le \beta \le \delta_k,$$

which shows that s^* is also feasible for $SUB(x_k, p_k; \delta_k)$. By (3), s_k obtained from the subproblem $SUB(x_k, p_k; \delta_k)$ satisfies

$$\begin{aligned} f(x_k) - m(x_k, p_k)(s_k) &\ge \tau[f(x_k) - m(x_k, p_k)(s_k^*)] \\ &\ge \tau[f(x_k) - m(x_k, p_k)(s^*)] \\ &> \tau\,\frac{\eta^*}{2}, \end{aligned}$$

where s_k^* is the exact solution of $SUB(x_k, p_k; \delta_k)$ and $k \in K'$. By (14), it follows that

$$f(x_k) - f(x_{k+1}) > c_0\,\tau\,\frac{\eta^*}{2} \tag{15}$$

for k large enough and $k \in K'$.

However, since the series with positive terms

$$\begin{aligned} \sum_{j=1}^{\infty}[f(x_{k_{i_j}}) - f(x_{k_{i_j}+1})] &\le \sum_{j=1}^{\infty}[f(x_{k_{i_j}}) - f(x_{k_{i_{j+1}}})] \\ &= f(x_{k_{i_1}}) - f(x^*) < +\infty, \end{aligned}$$

is convergent, we have

$$f(x_{k_{i_j}}) - f(x_{k_{i_j}+1}) \to 0, \quad as\ j \to \infty.$$

This contradicts (15) and completes the proof. □

We can also state our convergence theorem in the following manner.

Corollary 4.4 *Under Assumptions 2.1 through 2.5, if*

1. *the level set $L_0 = \{x \in R^n \mid f(x) \le f(x_0)\}$ is bounded, where x_0 is the starting point of the TR iteration,*

or

2. *the sequence $\{x_k\}$ generated by the basic TR algorithm is bounded,*

then the sequence $\{x_k\}$ has at least one accumulation point, and every accumulation point of $\{x_k\}$ is a stationary point of f.

5 Conclusions

In this paper we have identified five reasonable assumptions, *i.e.*, Assumptions 2.1 through 2.5, which have allowed us to produce a general global convergence theory for trust region methods for nonsmooth optimization. We have demonstrated that this theory can be viewed as a unified approach to convergence analysis by showing that a global convergence theory for each of four very distinct TR applications in the literature can be obtained as special cases of our general approach. In two of these applications we were forced to make stronger assumptions, but produced a stronger convergence theory.

In the cases studied in Section 3, the parameters in the TR local models could represent information related to first derivatives, second derivatives, or Lagrange multipliers. As a unified approach, we assumed the boundedness of these parameters in Assumption 2.5. This boundedness can be derived in many TR method applications from parameter updating strategies. The boundedness of the parameters is employed in the proof of Theorem 4.3 to guarantee the existence of a convergent subseqence. In particular applications, for example Powell (1984) and Yuan (1983), it may be possible to establish a convergence theory without assuming bounded parametric information. This boundedness assumption seems to be the price we had to pay for establishing a general theory.

Acknowledgements. The authors thank Mohammedi El Hallabi for sharing his understanding and thoughts on this subject with the authors in the formative stages of this research. His Ph.D. thesis [5] influenced the current research activity. They also thank Richard Byrd and Peppino Terpolilli for valuable interactions and discussions. Finally, they thank Nick Gould for suggestions that led to an improvement in the presentation.

References

[1] R.H. Byrd, R.B. Schnabel, and G.A. Shultz, "Approximate solution of the trust region problem by minimization over two-dimensional sub-

the trust region problem by minimization over two-dimensional subspaces", Mathematical Programming 40(1988)247-263.

[2] F.H. Clarke, Optimization and nonsmooth analysis (Canadian Math. Soc. Series of Monographs and Advanced Texts, John Wiley and Sons, 1983).

[3] J.E. Dennis Jr. and R.B. Schnabel, Numerical methods for unconstrained optimization and nonlinear equations (Prentice-Hall, Englewood Cliffs, New Jersey, 1983).

[4] I.S. Duff, J. Nocedal, and J.K. Reid, "The use of linear programming for the solution of sparse sets of nonlinear equations", SIAM Journal on Scientific and Statistical Computing 8(1987)99-108.

[5] M. El Hallabi, "A global convergence theory for arbitary norm trust region methods for nonlinear equations", Ph.D. Thesis, Rice University (Houston, TX, 1987).

[6] M. El Hallabi and R.A. Tapia, "A global convergence theory for arbitrary norm trust region methods for nonlinear equations", Technical Report 87-25, Department of Mathematical Sciences, Rice University (Houston TX, 1987).

[7] R. Fletcher, "A model algorithm for composite nondifferentiable optimization problem", Mathematical Programming Studies 17(1982)67-76.

[8] R. Fletcher, "An L_1-penalty method for nonlinear constraints", in: P. Boggs, R. Byrd and R. Schnabel, eds., Numerical optimization (SIAM, Philadelphia, 1984) pp. 26-40.

[9] R. Fletcher, Practical methods of optimization (John Wiley and Sons, Chichester, 1987).

[10] S.-B.B. Li, "Global convergence of trust region methods for minimizing a nondifferentiable function", Ph.D. Thesis, Rice University (Houston, TX, 1989).

[11] J.J. Moré, "Recent developments in algorithms and software for trust region methods", in; A. Bachern, M. Grotschel and B. Korte, eds., Mathematical programming: the state of the art (Spring Verlag, Berlin, 1983) pp. 258-287.

[12] J.M. Ortega and W.C. Rheinboldt, Iterative solution of nonlinear equations in several variables (Academic Press, New York, 1970).

[13] M.J.D. Powell, "General algorithms for discrete nonlinear approximation calculations", Report DAMTP 1983/NA2, Department of Applied Mathematics and Theoretical Physics, University of Cambridge (Cambridge, England, 1983).

[14] M.J.D. Powell, "On the global convergence of trust region algorithms for unconstrained minimization", Mathematical Programming 29(1984)297-303.

[15] R.T. Rockafellar, Convex analysis (Princeton Mathematics Services, Princeton University, Princeton, NJ, 1970).

[16] Y. Yuan, "Global convergence of trust region algorithms for nonsmooth optimization", Report DAMTP 1983/NA13, Department of Applied Mathematics and Theoretical Physics, University of Cambridge (Cambridge, England, 1983).

[17] J. Zhang, N.-H. Kim and L. Lasdon, "An improved successive linear programming algorithm", Management Science 31(1985)1312-1331.